

Gregor Neuböck

Alte Bücher – neue Leser/-innen

Ein Bericht aus der Digitalisierungswerkstätte der Oö. Landesbibliothek

Gemessen an der Anzahl der Datensätze im Bibliothekenverbund ist die Oö. Landesbibliothek zur zehntgrößten Büchersammlung in Österreich angewachsen. Als Landesbibliothek liegt ihr jedoch die Pflege der Sammlungen aus und über Oberösterreich besonders am Herzen („Obderennsia“). Mit dem Aufkommen der institutionellen Repositorien hat die Bibliothek begonnen, sich auch der Digitalisierung und Erschließung dieser urheberrechtsfreien Bücher anzunehmen. Es ist kein Zweifel, dass die Digitalisierung den Büchern ein „zweites Leben“ beschert hat. Die Bibliothek kann ihr Profil „schärfen“ und im Konzert der großen Informationsanbieter mit ihren unikalen Beständen mitmischen. Ein Nebeneffekt des weitgehend „inhouse“ organisierten Digitalisierungsprozesses ist der Aufbau von „Know-how“, der der Bibliothek eine informationstechnologische Kompetenz zurückgibt, die sie in Zeiten des „Outsourcings“ von technischen Informationsdienstleistungen schon stark aus der Hand gegeben hatte.

Die Oö. Landesbibliothek liegt im Zentrum von Linz an der Geschäftsstraße Landstraße in unmittelbarer Nähe zum neuen Musiktheater und bildet zusammen mit diesem und dem Wissensturm, einer gemeinsamen Dachmarke von Stadtbibliothek und Volkshochschule, einen neuen „Bildungsbezirk“ im Umfeld des Bahnhofsviertels. Sie ist eine klassische Universalbibliothek mit einer langen Tradition als älteste öffentliche Bibliothek des Landes Oberösterreich, über 100 Jahre als „Studienbibliothek“ und mit dem „Land Oberösterreich“ als Unterhaltsträger seit 1999. Ihrer Funktion nach ist sie Dienstleistungsbetrieb und Informationslieferant für die (vor)wissenschaftlichen Arbeiten der Mittelschüler, für das berufliche und private Lernen, sie ist aber auch Archivbibliothek für die Publikationen aus und über Oberösterreich, und als Bewahrerin des Büchererbes der aufgehobenen Klöster und Stifte ist sie auch eine museale Kultureinrichtung. Die Bibliothek hat aber auch regen Zulauf von Studentinnen und Studenten, die an anderen Studienorten eingeschrieben sind, die die Landesbibliothek als Quelle und Lernort für das Schreiben von Fach- und Diplomarbeiten nutzen.

Seit dem Beitritt der Landesbibliothek zum österreichischen Bibliothekenverbund im Jahr 2001 hat die Bibliothek einen stetig steigenden Zulauf von Leserinnen und Lesern quer durch alle Altersgruppen, weil ihre Bestände im Portal des Verbundes entsprechend abgebildet werden. Gemessen an der Anzahl der Datensätze im Bibliothekenverbund ist die Bibliothek zur zehntgrößten Büchersammlung in Österreich angewachsen und weist nahezu ebenso viele bibliographische Eintragungen auf wie die Universitätsbibliothek der Kepler-Universität. Neben der verbesserten Präsenz der Bestände im österreichweiten Datenpool brachte ein Zubau- und Sanierungsprojekt in zeitlicher Abstimmung mit

der Wahl von Linz als Kulturhauptstadt 2009 eine wesentlich attraktivere Präsentation der Bestände durch Freihandaufstellung, gut ausgestattete Lernplätze und automatisierte Ausleiheprozesse.

Als Landesbibliothek liegt ihr jedoch die Pflege der Sammlungen aus und über Oberösterreich besonders am Herzen („Obderennsia“), bedient sie doch gerade auch das Segment der Heimat- und Regionalforscher mit Literatur der historisch gewachsenen Sammlungen. Diese historischen Bestände mit einer Vielzahl von gemeinfreien Werken aus und über Oberösterreich wurden durch Konversion der Metadaten in den Verbundkatalog nach und nach im Portal suchbar gemacht und lösen somit die mittlerweile digitalisierten Zettelkataloge sukzessive ab.

Mit dem Aufkommen der institutionellen Repositorien hat die Bibliothek begonnen, sich auch der Digitalisierung und Erschließung dieser urheberrechtsfreien Bücher anzunehmen. Damit wollte man den stärker werdenden Bedürfnissen des Lesepublikums nach Volltexten einerseits entgegenkommen und andererseits die landeskundlichen Inhalte ortsunabhängig an den vernetzten Arbeitsplätzen der Benutzer auch außerhalb der Bibliothek und außerhalb Oberösterreichs verfügbar machen. Die Erwartung, wenig genutzte alte Bestände auf diese Weise wieder für historisch interessiertes Publikum zugänglich zu machen, hat sich mehr als erfüllt. Wie die Zugriffsstatistiken zeigen, erreichen die digitalisierten Bücher das Publikum weit über Oberösterreich hinaus, insbesondere in Osteuropa und im angloamerikanischen Sprachraum. Es ist kein Zweifel, dass die Digitalisierung den Büchern ein „zweites Leben“ beschert hat. Die Bibliothek kann ihr Profil „schärfen“ und im Konzert der großen Informationsanbieter mit ihren unikalen Beständen mitmischen. Ein Nebeneffekt des weitgehend „inhouse“ organisierten Digitalisierungsprozesses ist der Aufbau von „Know-how“, der der Bibliothek eine informationstechnologische Kompetenz zurückgibt, die sie in Zeiten des „Outsourcings“ von technischen Informationsdienstleistungen schon stark aus der Hand gegeben hatte.

Nachdem die Gebäudesanierung und der Zubau wesentliche Ressourcen für die Bibliotheksentwicklung gebunden hatten, war der Blick nach der Eröffnung bzw. Wiedereröffnung des Standortes wieder frei für anstehende bibliothekarische Entwicklungen. Bedingt durch den Umstand, dass Linz erst in den 60er Jahren des vorigen Jahrhunderts Universitätsstadt wurde, lagen bei der Studien- bzw. späteren Landesbibliothek die Sammlungsschwerpunkte immer schon bei der Literatur aus und über Oberösterreich und bei Beständen aus oberösterreichischen Klöstern, die im Zuge der Jesuitenauflösungen nicht nach Wien oder in andere Universitätsstädte gekommen sind.

1. Chronologie der Entwicklung der „Digitalen Landesbibliothek Oberösterreich“ (DLOÖ)

Ende 2009 wurde an der Oberösterreichischen Landesbibliothek begonnen, ein Digitalisierungsprojekt zu entwickeln. Zunächst setzten wir uns mit Standards im Bereich der Digitalisierung auseinander. Dazu zählten insbesondere Dateiformate für Bilder, Meta- und Strukturdaten sowie OCR-Daten. Zusätzlich wurden verschiedene Bibliotheken und Firmen kontaktiert und besucht, die schon Erfahrungen in diesem Bereich vorweisen konnten.

Im Jahr 2010 legten wir dann unsere Kriterien für eine geeignete Software fest und entschieden uns Ende 2010 für die Open Source-Software¹ Goobi. Mangels geeigneter Personalressourcen lagerten wir die technischen Arbeiten an Goobi, die zur Vorbereitung nötig waren, an die Firma Intranda in Göttingen aus. Im März 2011 begannen wir schließlich nach einer halbtägigen Einschulung mit der operativen Arbeit auf einem gehosteten² Webspaces, und schon kurze Zeit später erschienen unsere ersten urheberrechtsfreien Bücher im Internet.

In den ersten acht Monaten, der einem Probebetrieb gleichzusetzen war, scannten wir unsere Bücher auf einem Flachbettscanner. Dementsprechend digitalisierten wir in diesem Zeitraum nur kleinformatige Bücher, die in sehr gutem Zustand waren.

Mit der Anschaffung eines eigenen Servers inklusive Sicherungsservers sowie eines Auflichtscanners der Marke CopyBook A2+ gegen Ende des Jahres 2011 machten wir dann den nächsten großen Schritt, denn damit hatten wir nun alle notwendigen Komponenten im eigenen Haus.

Dieser Schritt war auch deshalb besonders wichtig, weil wir gleich anfangs die Entscheidung getroffen hatten, unsere Werke mit einer hohen Auflösung zu digitalisieren. Bis auf wenige Ausnahmen, wie z.B. großformatige Kartenbeilagen, scannen wir ausschließlich im Format 600 dpi³. Die mit den großen Datenmengen verbundenen Probleme des Datenuploads zum externen Host zwangen uns praktisch zu einer Inhouse-Lösung, denn schon kleine Bücher benötigen bei dieser Auflösung mehrere GB⁴ Speicherplatz.

¹ Open Source-Software bietet einen offenen Quelltext an, der verändert und frei genutzt werden darf. Lizenzrechtlich werden dabei die Lizenzbedingungen der Open Source Initiative anerkannt.

² Unter Webhoster versteht man Internetdienstleister, die in erster Linie Webspaces zur Verfügung stellen. Oft werden aber auch noch weitere Dienste wie z.B. E-Mail oder Datenbanken bereitgestellt.

³ dpi steht für dots per inch und ist ein Maß für die Punktdichte. Diese bestimmt entscheidend die Qualität von Bildern.

⁴ 1 GB entspricht dem Speicherplatzbedarf von 1 000 000 000 Byte.

Ab Dezember 2011 hatten wir dann ein kleines Digitalisierungs-LAN⁵ aufgebaut. Damit konnten die Bilder nun direkt vom Scanner in die Goobi-Vorgangsordner eingespielt werden. Zusätzlich wurde ein 27-Zoll- iMac angeschafft, der wegen seiner großen Bild diagonalen und der sehr guten Darstellung von Bildern zur Qualitätskontrolle verwendet wird.

Als zusätzliche Komponente, speziell für die Digitalisierung von Handschriften und Büchern mit einem kleinen Öffnungswinkel, kauften wir im Dezember 2012 von VESTIGIA – The Manuscript Research Centre of Graz University – den Traveller, einen mobilen Digitalisierungstisch, der eine besonders schonende Digitalisierung erlaubt.

Seit Herbst 2013 besitzen wir außerdem, sowohl für Goobi als auch für unseren Viewer⁶, einen eigenen physischen Server. Zusätzlich werden diese beiden Systeme über einen Sicherungsserver gesichert. Als Datensicherung setzen wir zwei große NAS⁷ mit Raid⁸ Level 5 ein, auf denen alle Daten zusätzlich doppelt gesichert werden. Gegen Stromausfälle wird das gesamte System mit zwei USV⁹ abgesichert.

Goobi und der Viewer werden seit ihrer Implementierung einer ständigen Weiterentwicklung bezüglich Anforderungen und Standards unterzogen. In enger Zusammenarbeit mit der Goobi-Community und unserer Betreuungsfirma werden die beiden Komponenten entsprechend den Marktanforderungen und technischen Änderungen angepasst.

2. Entscheidungskriterien für die Komponenten der DLOÖ

Aus Platzgründen werde ich hier nur auf die Entscheidungsgründe im Bereich der Digitalisierungssoftware und des Scanners eingehen und übergehe alle anderen Komponenten wie z.B. NAS, Server, USV usw.

Ein wesentlicher Grund für die Auswahl von Goobi als Digitalisierungsplattform war für uns die Plattformunabhängigkeit. Das bedeutet, dass man im Gegensatz zu vielen Kon-

⁵ LAN ist ein lokales Rechnernetzwerk, bei dem in unserem Fall Server, Scanner und einige ausgewählte Computer eingebunden sind.

⁶ Als Viewer bezeichnen wir hier unser Präsentationssystem, basierend auf einer Eigenentwicklung der Fa. IntraData.

⁷ NAS ist ein netzwerkgebundener Dateiserver, der in einem Rechnernetz Speicherkapazität zur Verfügung stellt.

⁸ RAID steht für redundante Anordnung unabhängiger Festplatten, RAID 5 ist eine Anordnung von Festplatten mit einem guten Kompromiss zwischen Datendurchsatz, Redundanz und Kosten.

⁹ USV ist eine unterbrechungsfreie Stromversorgung. Wird dazu eingesetzt, um bei Störungen im Stromnetz die Versorgung unserer Server zu garantieren.

kurrenzprodukten, die einen lokal installierten Software-Client¹⁰ benötigen, nur einen Internetzugang und einen Browser benötigt, um im Produktionssystem arbeiten zu können. Als weitere Gründe sind noch das Fehlen von Lizenzgebühren zu nennen, was die laufenden Kosten massiv senkt, und die gute Konfigurierbarkeit des Systems, wodurch sich Goobi für uns zu einer Art Schweizermesser der Digitalisierung entwickelt hat.

Heute digitalisieren wir unterschiedlichste Werktypen wie z.B. Monographien, Zeitungen, Bandserien, Handschriften, Karten, Urkunden und Inkunabeln. Zur Einbindung externer Dienstleister, z.B. zur Digitalisierung sensibler Handschriften, wird von uns ein spezifischer Benutzerzugang eingerichtet. Die Bilder werden dann vom Dienstleister über das Internet in Goobi hochgeladen, noch bevor die Handschriften wieder im eigenen Tresor zurück sind. Es kommt auch vor, dass wir für Wissenschaftler einen exklusiven Zugang zu einer bestimmten Handschrift einrichten, damit besonderes Expertenwissen in deren Beschreibung einfließen kann. Der Zugang kann in wenigen Minuten exklusiv eingerichtet werden, und der Wissenschaftler kann nach kurzer Einschulung von seinem eigenen Computer, unabhängig von Zeit und Raum, an der Handschrift arbeiten. Dadurch profitieren nicht nur wir durch die bessere Erschließung unserer Handschriften, sondern auch die Handschriftenforscher weltweit.

Da wir ein sehr kleines Team von teilweise wechselnden Personen sind, ist die Browserbasiertheit sowie die Möglichkeit zum Anlegen unterschiedlicher Benutzergruppen und Benutzer ohne Lizenzgebühren oder andere Einschränkungen von sehr großem Vorteil.

Schlussendlich spielte auch die große Community, von der Goobi mittlerweile eingesetzt und weiterentwickelt wird, eine Rolle bei unserer Entscheidung. Denn diese stellt de facto eine Garantie für eine nachhaltige Entwicklung der Software dar, die zusätzlich auch noch durch den Verein „Goobi DIGITALISIEREN IM VEREIN“¹¹ rechtlich auf soliden Beinen steht. Im diesem Verein sind viele verschiedene Bibliotheken vertreten, die sich die – koordinierte – Weiterentwicklung von Goobi zur Aufgabe gemacht haben.

Für die Entscheidung betreffend den Auflichtscanner waren folgende Punkte maßgeblich: Wir forderten 600 dpi optischer Auflösung über eine Fläche von A2 in Farbe, eine Anforderung, die nur von wenigen Herstellern erfüllt werden konnte. Ausschlaggebend für das CopyBook von i2s war letztlich die Geräuschentwicklung und der Preis, da der Scanner in einer Büroumgebung mit mehreren Mitarbeitern eingesetzt wird. Dadurch,

¹⁰ Software-Client ist ein Computerprogramm, das benötigt wird, um auf das Produktionssystem zugreifen zu können.

¹¹ Goobi. Digitalisieren im Verein e. V. <http://www.goobi.org/>

dass dieser Scanner mit einer Kamera arbeitet und kein Zeilenscanner ist, hat er kaum bewegliche Teile, damit einen geringeren Geräuschpegel als diese und weniger mechanische Teile, die sich abnutzen können.

3. Der Digitalisierungsworkflow der DLOÖ

Goobi bietet die Möglichkeit, eine beliebige Anzahl von verschiedenen Arbeitsschritten miteinander zu verknüpfen. Im Projekt der Oberösterreichischen Landesbibliothek werden derzeit folgende Schritte im Digitalisierungsworkflow eingesetzt: bibliographische Aufnahme, Scannen, Qualitätskontrolle, automatischer Kopier- und Tiffheaderschritt, Struktur- und Metadaten, OCR¹², URN¹³-Generierung, Export in DMS¹⁴, URN-Eintragung und Archivierung. Die einzelnen Schritte können entweder automatisch oder manuell ablaufen und optional einer Validierung unterzogen werden. Der jeweils nächste Schritt kann erst nach erfolgreichem Abschluss des Vorgängerschrittes begonnen werden. Den einzelnen Arbeitsschritten können Benutzergruppen bzw. Benutzer zugeordnet werden. Man spricht in diesem Zusammenhang vom Prinzip des aufgeräumten Schreibtisches. Jeder Benutzer sieht nur die ihm zugeordneten Arbeitsschritte. Der Scanoperator sieht in seinem Workflow ausschließlich diejenigen Vorgänge, bei denen der Arbeitsschritt Scannen ansteht. Sobald dieser erfolgreich beendet ist, verschwindet der betreffende Vorgang für den Scanoperator. Das Gleiche gilt auch für den Metadatenbearbeiter, der die inhaltliche Strukturierung und Erschließung des Dokuments vornimmt. Hat dieser seine Arbeit erfolgreich abgeschlossen, verschwindet der Vorgang aus seinem Arbeitsbereich.

Da bei den einzelnen Schritten Fehler passieren können, hat Goobi ein internes Kommunikationssystem, mit dem ein Vorgang von einem Bearbeiter an einen vorhergegangenen Bearbeitungsschritt zurückgeschickt werden kann. Wenn z.B. der Metadatenbearbeiter bemerkt, dass Bilder nicht korrekt gescannt wurden oder Seiten fehlen, kann er den Vorgang an den Scanoperator zurückschicken, gleichzeitig mit einer Nachricht, die es diesem erleichtert, den Fehler zu entdecken und zu korrigieren.

¹² OCR steht für die automatisierte Texterkennung innerhalb von Bildern.

¹³ URN ist ein eindeutiger und dauerhafter Name für Ressourcen. Mit diesem wird der Fehler 404 vermieden, da einer Ressource immer nur genau ein persistenter Identifier zugeordnet wird.

¹⁴ DMS: Abkürzung für Dokumentenmanagement-System, im Text auch als Viewer bezeichnet

Bibliographische Aufnahme:

Die Daten werden direkt aus der Aleph-Datenbank über die AC-Nummer via Schnittstelle Z39.50 übernommen. Bei Zeitschriftenbänden kann in Goobi eine Vorlage angelegt werden. Bei den einzelnen Bänden werden dann nur mehr die Bandangaben ergänzt. Unsere Handschriften befinden sich nicht in der Aleph-Datenbank. In diesem Fall wurde unser Handschriftenkatalog (PDF) geparkt und durch die Eingabe der Handschriften-Signatur werden die entsprechenden Daten einer Handschrift in Goobi importiert.

Scannen:

Bücher werden bei uns ausnahmslos mit 600 dpi in Farbe gescannt. Entsprechende Datenmengen können nur dadurch bewältigt werden, weil unsere Server in einem kleinen LAN stehen und über eine Gigabit-Anbindung des Scanners die Dateien in Goobi eingespielt werden.

Qualitätskontrolle:

Unsere Mitarbeiter und Mitarbeiterinnen kontrollieren Vollständigkeit und Schärfe der Bilder auf einem iMAC mit 12 GB RAM. Hier kann bei Bedarf auch das schon zuvor erwähnte Rückmeldungssystem von Goobi eingesetzt werden. Sollten Fehler auftreten, so können diese in einem Textfeld beschrieben und an die betroffene Stelle (Scannen, Qualitätskontrolle ...) rückgemeldet werden.

Automatischer Kopier- und Tiffbeheaderschritt:

In diesem Schritt werden die vorhandenen Mastertiffs in einen neuen Ordner kopiert und einer JPEG-Komprimierung unterzogen (ca. 15% Speicherplatz der Ausgangsdatei). Der Ordner mit den Mastertiffs wird später (siehe Schritt Archivierung) auf eine NAS verschoben, um Speicherplatz am Server zu sparen.

OCR-Volltexterfassung:

Alle unsere Bücher, ob Antiqua- oder Frakturschrift, werden automatisiert über Abby Recognition-Server im Volltext erfasst. In den nächsten Wochen wird zusätzlich ein Web2.0-Crowdsourcing-Tool zur manuellen Verbesserung der OCR-Texte implementiert.

Struktur- und Metadatenerfassung:

Der Metadateneditor in Goobi gliedert sich in die drei Hauptschritte Paginierung, Strukturdatenerfassung und Metadateneingabe.

Bei der Paginierung werden die Images den Seiten zugeordnet. Es kann arabische oder römische Paginierung, unnummeriert und Freitexteingabe gewählt werden.

Bei der Strukturdatenvergabe wird grundsätzlich, sofern vorhanden, das Inhaltsverzeichnis abgebildet. Zusätzlich werden alle Bilder, Tabellen, Einbände, Titelblätter, Anmerkungen, Gedichte usw. erfasst. Die bei uns verwendeten Strukturelemente verwenden die Standards der Library of Congress, um die Validität der ausgegebenen METS-Datei zu garantieren.

Bei der Metadateneingabe werden dann die zuvor erstellten Strukturelemente mit Inhalt befüllt.

URN-Generierung:

Unsere URNs werden laut OBVSG-Policy automatisiert vergeben (z.B. urn:nbn:AT-OOeLB-9373) und durch den OBVSG-Resolver (<https://resolver.obvsg.at/>) aufgelöst. Zusätzlich kann Goobi auch seine eigenen URNs auflösen.

Export in DMS:

In diesem Schritt werden alle Bilder, Meta- und Strukturdaten sowie eventuell vorhandene OCR-Ergebnisse automatisch nach Abschluss des vorhergehenden Schrittes in den Viewer kopiert.

In unserem Fall sind Viewer und Produktionssystem auf zwei physischen Maschinen getrennt installiert, um jegliche Beeinflussung durch gegenseitige Lastzugriffe zu vermeiden.

Goobi kann grundsätzlich mit einer Vielzahl an unterschiedlichen Viewern zusammenarbeiten (Typo3, Intranda Viewer, MyCoRe, DigiTool, DFG-Viewer).

URN-Eintragung:

Derzeit werden noch alle unsere URNs an den OBVSG-Resolver gemeldet, und die URN wird manuell im Katalog eingetragen.

Unser System liefert allerdings jetzt schon alle Voraussetzungen, diese Daten automatisiert über OAI-Schnittstelle (<http://digi.landesbibliothek.at/viewer/oai/>) zur Verfügung zu stellen.

Archivierung:

Alle unsere Daten (Meta- und Strukturdaten, OCR-Ergebnisse) werden auf eine NAS kopiert, die Mastertiffs werden verschoben. Die Dateien werden als TAR-Archiv mit einem Datumsstempel versehen. Die nachträgliche individuelle Archivierung von Metadaten oder OCR-Ergebnissen ist jederzeit möglich und wird durch den Datumsstempel sofort als die aktuelle Version kenntlich. Eine zusätzliche Sicherung auf eine weitere NAS wird in den Tresorraum verbracht.

Was ich als Administrator an der Arbeit mit Goobi am meisten schätze, ist die große Flexibilität, die mir diese Software gibt. Wird ein Buch oder eine Zeitschrift oder was auch immer zur Digitalisierung angefragt, so kann ich jeden Workflowschritt selber bestimmen, und manchmal dauert es nur wenige Stunden, bis ein angefragtes Buch online ist.

4. Womit kann man im Konzert der Großen bestehen?

Wie kann man als mittelgroße oder kleine Regionalbibliothek mit den großen Bibliotheken mithalten oder vielleicht sogar noch besser sein?

Der Vorteil einer kleineren Bibliothek gegenüber den großen besteht nicht zuletzt darin, dass sie – weil Massendigitalisierung ohnehin eine Domäne der großen Bibliotheken ist – selektiv vorgehen kann und damit verbunden auch hohe Qualitätsstandards festlegen kann.

Durch die riesige Menge an Digitalisaten werden große Bibliotheken in der Regel zu einem niedrigeren Qualitätsniveau gezwungen, sowohl was die Qualität der Scans als auch was die Erschließungstiefe betrifft. Genau dort setzen wir als Landesbibliothek an. Alle unsere Werke werden mit einer Auflösung von 600 dpi Farbe gescannt (was sich wiederum in der Qualität des OCR niederschlägt), nur großformatige Karten werden mit 400 dpi gescannt. Bei der Erfassung der Struktur- und Metadaten bilden wir das Inhaltsverzeichnis ab, zusätzlich werden aber auch noch andere Elemente wie Illustrationen, Tabellen, Gedichte, Noten etc. als Strukturelemente erfasst.

Das Inhaltsverzeichnis und alle weiteren Strukturelemente werden nach dem Export im Viewer als verlinkte Navigation links vom Bild dargestellt. Alle erfassten Teile können selektiv in Faksimile-Qualität als PDF heruntergeladen werden. Die intensive Erschließung durch Struktur- und Metadaten wirkt sich letztlich auch auf das Ranking unserer Digitalisate bei der Suche in Google aus, wo diese regelmäßig auf vorderen Plätzen auftauchen.

Auch bei uns läuft jedes Buch, gleich ob es in Antiqua- oder Frakturschrift gedruckt wurde, durch einen automatisierten OCR-Schritt. Über diesen Volltext werden unsere

Bücher von Suchmaschinen leichter aufgefunden. Die Verfügbarkeit von Volltext wird gerne als Argument gegen die hohe Erfassungstiefe eingesetzt, so als sei diese unnütz oder eine unnötige Fleißaufgabe. Abgesehen von der Zuverlässigkeit der Buchstabenerkennung durch OCR ist es aber für die Orientierung im Text dennoch wichtig, bei einem digitalisierten Buch auch die Struktur – das Inhaltsverzeichnis – abzubilden, um einen Überblick zu haben und sich im Text vernünftig bewegen zu können.

Neben der Qualität der Meta- und Strukturdaten ist auch die Bildqualität entscheidend. Insbesondere bei unseren Handschriften (<http://digi.landesbibliothek.at/viewer/fullscreen/471/28/>) kommt dies ganz deutlich zur Geltung. Hier werden über die Zoomfunktion und die Vollbildanzeige Details sichtbar, für die man beim Original eine Lupe bräuchte. Auch bei großformatigen Karten und Illustrationswerken (<http://digi.landesbibliothek.at/viewer/fullscreen/AC00969620/13/>) oder Atlanten (<http://digi.landesbibliothek.at/viewer/fullscreen/AC05371191/265/>) ist dies der Fall.

Als weiteren Service bieten wir auch das Vorziehen von angefragten Büchern im Digitalisierungsprozess. Damit diese rasch verfügbar sind, werden die Bücher nur mit den bibliographischen Metadaten versehen und sofort exportiert. Erst in weiterer Folge werden Inhaltsverzeichnisse strukturell abgebildet und später nachträglich exportiert. Damit können digitalisierte Bücher rasch zur Benützung freigegeben werden, und trotzdem kann der Anspruch auf hohe Qualität bei der Erfassung der Struktur- und Metadaten aufrecht bleiben.

Die konsequente Weiterentwicklung unseres Portals wird einerseits für Goobi durch die Community, andererseits für den Viewer durch eine intensive und laufende Zusammenarbeit mit unserer Betreuungsfirma Intranda gewährleistet.

Zuletzt haben wir eine neue Suche (<http://digi.landesbibliothek.at/viewer/search/>) entwickelt, die neben der Standardsuche eine differenzierte Suche, eine Jahressuche und eine Jahresansicht bietet.

Bei Handschriften haben wir ein Beschreibungsfeld realisiert, in dem HTML-Tags für Verlinkungen, Listen, Hervorhebungen usw. verwendet werden können. So können mit der Zeit auf wissenschaftlicher Basis sehr umfangreiche und detaillierte Handschriftenbeschreibungen entstehen. Der Zugriff auf dieses Feld ist nur über eine exklusive Benützerberechtigung möglich.

Benutzer mit einem OpenID-Konto¹⁵ haben die Möglichkeit, sich eigene Sammlungen (Bücherregale) zusammenzustellen. So können je nach eigenem Interesse bestimmte Werke rasch auffindbar gemacht werden.

Die Normdatenanbindung direkt aus Goobi heraus wurde gerade realisiert. Nun werden die Normdatenanbindungen schon bei der „Bibliographischen Aufnahme“ in einen Vorgang importiert. Ein Tool mit Crowdsourcing¹⁶-Funktion zur Verbesserung der OCR-Ergebnisse wird in den nächsten Wochen freigeschaltet.

5. Zahlen und Fakten

Neben dem fallweisen Einsatz von Praktikanten als Scanoperatoren arbeiten je nach Verfügbarkeit ca. 1 - 1,5 Vollzeitbeschäftigte an der DLOÖ.

Die Oberösterreichische Landesbibliothek besitzt ca. 40.000 Drucke des 16. - 18. Jahrhunderts, 800 Inkunabeln und ca. 1.000 Handschriften (davon ca. 350 mittelalterliche). Neben diesen Schätzen sind es vor allem die sogenannten Obderennsia, die Literatur aus und über Oberösterreich, die im Zentrum unserer Digitalisierungsbemühungen steht. Landeskundliche Literatur, die online gestellt wurde, wird plötzlich wieder gelesen, nachdem manche Bücher vielleicht 80 Jahre oder länger nicht ausgehoben wurden! Seit 2012 digitalisieren wir zusätzlich schwerpunktmäßig Literatur zum Thema Erster Weltkrieg. Diese Entscheidung hat unsere Zugriffszahlen stark erhöht, was ich angesichts der häufig schon digitalisierten Werke auf die unter Punkt 4 angeführten Fakten zurückführen.

Die Oberösterreichische Landesbibliothek setzt die Digitalisierung auch zum Bestandschutz ein: Werke, die digitalisiert sind, dürfen im Regelfall nicht mehr entlehnt werden. Gerade Werke um die Wende vom 19. zum 20. Jahrhundert sind häufig in einem schlechten physischen Zustand, weitere Beschädigungen durch Benützung können durch die Digitalisierung vermieden werden.

Zum Schluss noch einige Bemerkungen zur Entwicklung unserer Zugriffe: Seit Dezember 2011 haben wir eine statistische Auswertung zur Verfügung. Damals, so sei erwähnt, hatten wir nur wenige Dutzend Werke online. Heute sind mehr als 2900 Werke online, die bis auf die Handschriften durchgängig im Volltext vorliegen.

¹⁵ OpenID ist ein dezentrales Authentifizierungssystem das nur eine einmalige OpenID von einem Anbieter benötigt. Danach kann man sich bei verschiedensten Systemen ohne Passwort und Benutzernamen anmelden.

¹⁶ Beim Crowdsourcing werden ursprünglich interne Arbeiten an die Internetcommunity ausgelagert (z.B. die Verbesserung von OCR-Ergebnissen).

Im Dezember 2011 hatten wir ca. 20 - 40 Zugriffe pro Tag bzw. 250 Seitenansichten. Im August 2014 waren es ca. 250 - 300 Besucher pro Tag oder bis zu 5.000 Seitenansichten aus mehr als 100 Ländern weltweit, und die Entwicklung setzt sich weiter positiv fort.

Zusätzlich sind wir über unsere OAI-Schnittstelle an verschiedene Portale wie z.B. „Zentrales Verzeichnis Digitalisierter Drucke“¹⁷ oder „Europeana“¹⁸ angebunden.

Fazit und Ausblick:

Eine Folge des Projektes ist, dass durch die nahezu vollständige Projektentwicklung im Hause und dadurch, dass der gesamte Workflow „inhouse“ bearbeitet wird, ein hohes Know-how erworben wurde. Während man durch „Outsourcing“ zwar Ressourcen spart, aber auch „Know-how“ verliert, ist es mit der DLOÖ gelungen, informations-technologische Lösungen selber weiterzuentwickeln bzw. einfach „Learning by Doing“ zu betreiben.

Besonders überraschend für uns ist auch die Resonanz auf das digitale Angebot außerhalb der Region und außerhalb des deutschen Sprachraumes. Insbesondere zu den Digitalisaten rund um Publikationen aus der Zeit des Ersten Weltkrieges erreichen uns Rückmeldungen aus der ganzen Welt, explizit aus dem angloamerikanischen Raum und aus Osteuropa. Die Resonanz und internationale Präsenz ist durchaus erwünscht, nahezu jede Veröffentlichung wird aber auch von den diversen Communities wahrgenommen und im Einzelfall per Rückmeldung oder in Foren kommentiert.

Die Oö. Landesbibliothek hat mit ihrem institutionellen Repositorium jedenfalls ein eigenständiges „Produkt“ entwickelt, das ihr im Umfeld der oberösterreichischen wissenschaftlichen Bibliotheken ein Alleinstellungsmerkmal ermöglicht.

Quellen:

Neuböck, Gregor Oberösterreichische Landesbibliothek: Neues Leben für alte Bücher – oder wie man ein Digitalisierungsprojekt auf Schiene bringt! *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare*, 2013, vol. 66, n. 1, pp. 179-191.

<http://de.wikipedia.org/wiki/Hosting>

<http://de.wikipedia.org/wiki/Punktdichte>

http://de.wikipedia.org/wiki/Gigabyte_%28Einheit%29

http://de.wikipedia.org/wiki/Local_Area_Network

http://de.wikipedia.org/wiki/Network_Attached_Storage

¹⁷ ZVDD [http://www.zvdd.de/dms/browsesammlungen/?tx_goobit3_search\[extquery\]=iswork%3A1&dc=zvdd.digi.landesbibliothek.at](http://www.zvdd.de/dms/browsesammlungen/?tx_goobit3_search[extquery]=iswork%3A1&dc=zvdd.digi.landesbibliothek.at)

¹⁸ <http://www.europeana.eu>

<http://de.wikipedia.org/wiki/RAID>

http://de.wikipedia.org/wiki/RAID#RAID_5:_Leistung_.2B_Parit.C3.A4t.2C_BlockLevel_Striping_mit_verteilter_Parit.C3.A4tsinformation

<http://de.wikipedia.org/wiki/Texterkennung>

<http://de.wikipedia.org/wiki/OpenID>

<http://de.wikipedia.org/wiki/Crowdsourcing>

http://de.wikipedia.org/wiki/Open_Source